Machine Learning Engineer Nanodegree

Capstone Project Report

Dog Breed Classifier

Manish Chandra

25 September, 2020

Contents

1 Definition

1.1 Domain Background

There are millions of dog owners throughout the world. The world has an estimated dog population of around 900 million [1]. With over 300 breeds of dogs [2], many dog owners might not be aware of the different breeds in existence or the breed(s) of their own dog(s). The importance of breed awareness could arise when considering buying a dog, its food or medical visits to the veterinary physician.

Recognising dogs according to their respective breed is a challenging task even for humans. There are over 300 breeds in existence which are grouped into 10 distinct groups according to physical characteristics. This shows that such classification is a complex task, as remembering all the breeds and distinguishing between similar breeds cannot be easily done by humans. This is a multi-class classification problem where we can use a supervised learning approach. Image classification refers to a task in computer vision of classifying an image according to its visual content. A convolutional neural network (CNN) has shown great success in the image domain. CNNs are a specialized type of neural network for processing data that have the structure of a grid.

1.2 Problem Statement

The task is to build a pipeline to process real world, user-supplied images. Given an image of a dog, the algorithm should give an estimate of the dog's breed. And when the image is of a human, the algorithm should give an estimate of a dog's breed that resembles the human in the image. If neither a dog nor a human is detected in the image, the algorithm should output an error message. Furthermore, the model should achieve 60 percent or greater accuracy.

1.3 Metrics

The metric used to compare the performance of different models is accuracy. This is so because the problem at hand is a classification task, where the model should classify the images accurately. Accuracy can be defined as the percentage of correct predictions out of all predictions.

2 Analysis

2.1 Data Exploration

The dataset has been provided by Udacity. The dog dataset [3] has 8351 dog images and the human dataset [4] has 13233 human images.

2.1.1 Dog Images Dataset

The dataset contains 8,351 dog images from 133 breeds. It is subdivided into training, validation and test data. There are 6,680 images to train the model, 835 images for validation and 836 images for testing. Two sample images from the dataset are shown in figures 1 and 2:

2.1.2 Human Images Dataset

This dataset contains 13233 images of humans. All the images are of size 250 × 250. Images have different backgrounds. These images are used to test the performance of the Human Face Detector.

Figure 1: Akita

Figure 2: German Shepherd

Figure 3: Human 1

Figure 4: Human 2

Two sample images from the dataset are shown in figures 3 and 4:

2.2 Data Visualization

In this section, we will check if the dog dataset suffers from data imbalance problem. The count of images of each dog breed across training, valid and test sets is shown in figure 5.

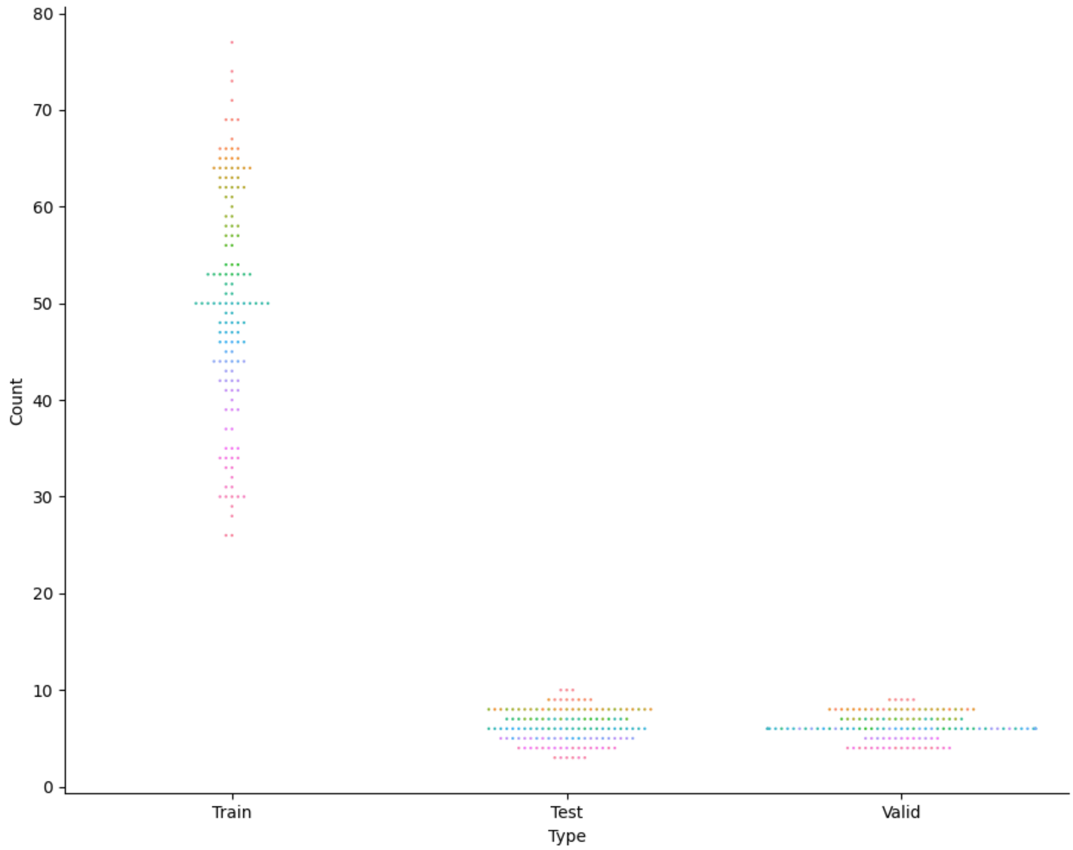

Figure 5: Dog Breed Counts

The images in test and validation sets are very well distributed, as is evident from the small and tight regions for the image counts shown in the figure 5. However, it is quite evident from the figure that the training set suffers from data imbalance.

3 Methodology

3.1 Preprocessing

It has already been pointed out that the images in the dog dataset have different sizes. Therefore, all the images in the validation and test sets are resized to 224×224 and then are normalized. In the training set, data augmentation is performed by random horizontal flip and random rotation of 15 degrees. All the images are cropped to 224×224 by randomly resizing and finally, are normalized.

3.2 Implementation

3.2.1 Human Face Detector

OpenCV's implementation of Haar feature-based cascade classifiers is used to detect human faces in the user-supplied images. The images are converted to greyscale before being passed to a face detector. The face detector is tested with 100 images each from the human and dog datasets. The face detector detects human face in 99% of the human images and 13% of the dog images.

3.2.2 Dog Detector

A pre-trained (trained on ImageNet) VGG-16 model is used to detect dogs in images. ImageNet has more than 10 million URLs where each URL links to an image containing an object from one of 1000 categories. Dog breeds occur consecutively on a dictionary from ImageNet from keys 151 to 268 inclusively, i.e., from 'Chihuahua' to 'Mexican hairless'. Therefore, if the VGG-16 model returns the index in this range, a dog is detected in the image.

3.2.3 Dog Breed Classifier from Scratch

The CNN model created from scratch must have accuracy of at least 10%. This can confirm that the model is working, because a random guess will provide a correct answer roughly 1 in 133 times, which corresponds to an accuracy of less than 1%. This CNN model created from scratch acts as the benchmark model. After some tuning, the architecture that worked quite well is as follows:

- Convolutional Layer with 3 input channels and 36 output channels, kernel size 3, Relu activation
- Max pooling layer with kernel size 2
- Convolutional Layer with 36 input channels and 64 output channels, kernel size 3, Relu activation
- Max pooling layer with kernel size 2
- Convolutional Layer with 64 input channels and 128 output channels, kernel size 3, Relu activation
- Max pooling layer with kernel size 2
- Flattenning layer to convert the pooled feature maps to a 1D vector
- Dropout with a probability of 0.25
- Fully connected layer with 512 neurons, batch normalization and Relu activation
- Dropout with a probability of 0.25
- Fully connected layer with 256 neurons, batch normalization and Relu activation
- Dropout with a probability of 0.25
- Fully connected layer with 133 neurons, Relu activation

A learning rate of 0.03 is used and training continues for 40 epochs. Cross entropy is used as the loss function and Stochastic Gradient Descent is used as an optimizer. The plot in figure 6 shows the training and validation losses over the epochs during training. The model corresponding to the epoch with lowest validation loss is saved.

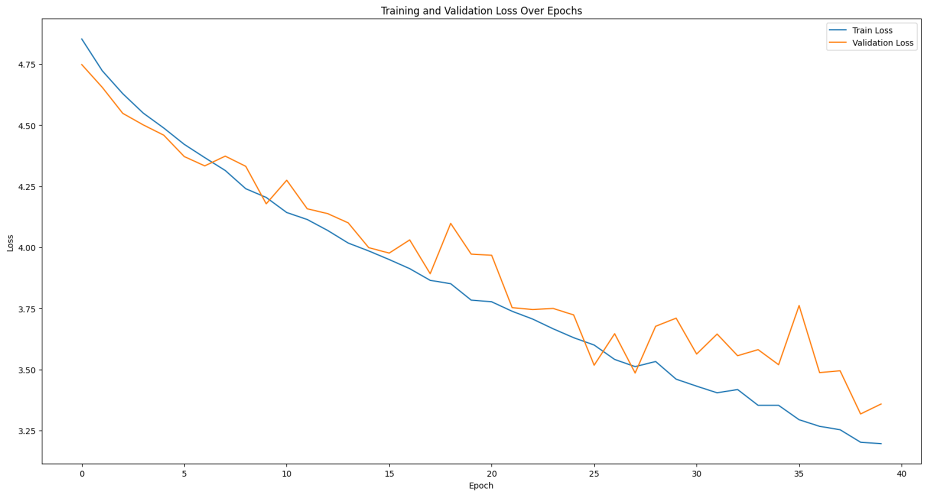

Figure 6: Training and Validation losses over epochs for the classifier built from scratch

3.2.4 Dog Breed Classifier using Transfer Learning

The CNN created from scratch has an accuracy of 24% on the test set. The performance can be sufficiently improved by using transfer learning. The pretrained model chosen for our purpose is the Resnet-152. The choice is based on the performance [5] of different architectures on ImageNet. Its feature extraction part is frozen, but its classifier part is modified to suit our problem. This is done by changing the number of output neurons to 133 (one for each dog breed). This modified classifier part is then trained.

A learning rate of 0.03 is used and training continues for 10 epochs. Cross entropy is used as the loss function and Stochastic Gradient Descent is used as an optimizer. The plot in figure 7 shows the training and validation losses over the epochs during training. The model corresponding to the epoch with lowest validation loss is saved.

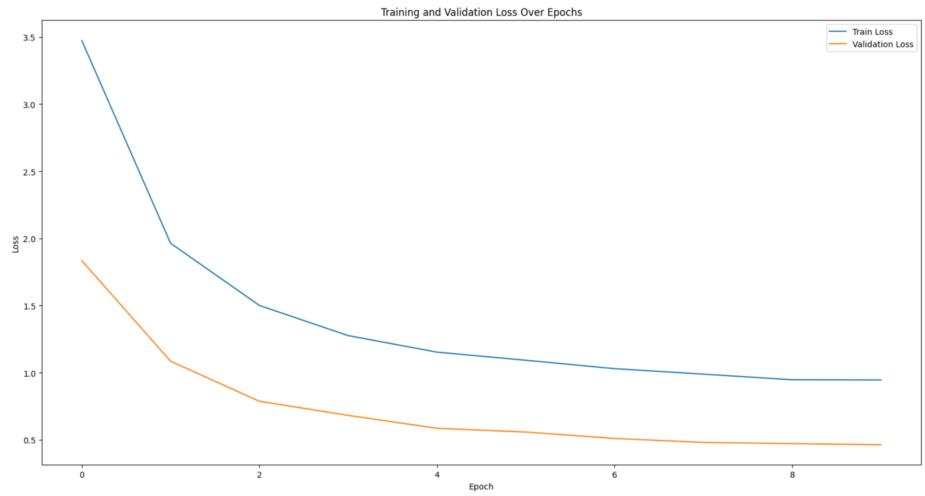

Figure 7: Training and Validation losses over epochs for the classifier using transfer learning

Figure 8: Result 1

Figure 9: Result 2

4 Results

The dog classifier from scratch achieved a test set accuracy of 24% (203/836) and that using transfer learning achieved 85% (714/836). Figures 8 to 13 show the predictions of the final amalgamated algorithm on the images that the models didn't see during training:

5 Conclusion

The results looks quite promising. However, the model made a mistake when fed with an image of a cat, predicting it to be human. The possible points of improvement are as follows:

- The first thing I would improve would be the face detector algorithm. I would build a new neural network using transfer learning for that.
- The second thing would be to improve the performance of the dog breed predictor by increasing the number of training epochs and tuning the hyperparameters further.
- The dog image dataset contains only 133 dog breeds. The FCI [6] currently recognizes 353 dog breeds. If this diversity were represented in the dataset, the algorithm would be able to provide an estimate for all these dog breeds.

```
Hello, dog!
You look like a ...
Pembroke welsh corgi
```

Figure 10: Result 3

```
Hello, human!
You look like a ...
Maltese
```

Figure 11: Result 4

```
Hello, human!
You look like a ...
Japanese chin
```

Figure 12: Result 5

```
Invalid image. Only human and dog images are accepted!
```

Figure 13: Result 6

Bibliography

[1] *Wikipedia*. URL: https://en.wikipedia.org/wiki/Free-ranging_dog (visited on 25/09/2020).

[2] *FCI Breeds Nomenclature*. URL: http://www.fci.be/en/Nomenclature/ (visited on 25/09/2020).

[3] Udacity. *Dog Images Dataset*. URL: https://s3-us-west-1.amazonaws.com/udacity-aind/dog-project/dogImages.zip (visited on 25/09/2020).

[4] Udacity. *Human Images Dataset*. URL: https://s3-us-west-1.amazonaws.com/udacity-aind/dog-project/lfw.zip (visited on 25/09/2020).

[5] Pytorch. *Pre-trained Models*. URL: https://pytorch.org/docs/stable/torchvision/models.html (visited on 25/09/2020).

[6] FCI. *For Pedigree Dogs Worldwide*. URL: http://www.fci.be/en/Presentation-of-our-organisation-4.html (visited on 25/09/2020).